



A population-based experimental model for protein evolution: Effects of mutation rate and selection stringency on evolutionary outcomes

Citation

Leconte, Aaron M., Bryan C. Dickinson, David D. Yang, Irene A. Chen, Benjamin Allen, and David R. Liu. 2013. A population-based experimental model for protein evolution: Effects of mutation rate and selection stringency on evolutionary outcomes. *Biochemistry* 52, no. 8: 1490–1499. doi:10.1021/bi3016185.

Published Version

doi:10.1021/bi3016185

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:27409156>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Published in final edited form as:

Biochemistry. 2013 February 26; 52(8): 1490–1499. doi:10.1021/bi3016185.

A population-based experimental model for protein evolution: effects of mutation rate and selection stringency on evolutionary outcomes

Aaron M. Leconte^{1,†,‡}, Bryan C. Dickinson^{1,‡}, David D. Yang¹, Irene A. Chen², Benjamin Allen^{3,4}, and David R. Liu^{1,*}

¹Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts, 02138

²Department of Chemistry and Biochemistry, University of California, Santa Barbara, 93106

³Program for Evolutionary Dynamics, Harvard University, Cambridge, Massachusetts, 02138

⁴Department of Mathematics, Emmanuel College, Boston, Massachusetts, 02115

Abstract

Protein evolution is a critical component of organismal evolution and a valuable method for the generation of useful molecules in the laboratory. Few studies, however, have experimentally characterized how fundamental parameters influence protein evolution outcomes over long evolutionary trajectories or multiple replicates. In this work, we applied phage-assisted continuous evolution (PACE) as an experimental platform to study evolving protein populations over hundreds of rounds of evolution. We varied evolutionary conditions as T7 RNA polymerase evolved to recognize the T3 promoter DNA sequence and characterized how specific combinations of both mutation rate and selection stringency reproducibly result in different evolutionary outcomes. We observed significant and dramatic increases in the activity of the evolved RNA polymerase variants on the desired target promoter after 96 hours of selection, confirming positive selection occurred under all conditions. We used high-throughput sequencing to quantitatively define convergent genetic solutions, including mutational “signatures” and non-signature mutations that map to specific regions of protein sequence. These findings illuminate key determinants of evolutionary outcomes, inform the design of future protein evolution experiments, and demonstrate the value of PACE as a method to study protein evolution.

While evolution plays an essential role both in shaping the natural world and in the development of valuable therapeutics, materials, and research tools^{1–6}, the determinants of evolutionary outcomes over long time courses both in nature and in the laboratory remain largely unexplored by systematic experimentation. Experimental efforts to understand protein evolution have largely relied on the reconstruction of presumed evolutionary intermediates^{7–10} or on experimental evolution over modest numbers of rounds of evolution

*Corresponding Author: Telephone: 617-496-1067. drliu@fas.harvard.edu.

[†]Present address

W.M. Keck Science Department of Claremont McKenna, Pitzer, and Scripps Colleges, Claremont, CA 91711

[‡]Author Contributions

These authors contributed equally to this manuscript.

The authors declare no competing financial interests

Supporting Information available

Supplementary Methods, Supplementary Discussion, Figures S1–S8, and Tables S1–S5. This material is available free of charge via the Internet at <http://pubs.acs.org>.

(typically fewer than ten)^{11–15}. The time-consuming nature of traditional directed evolution methods have made challenging the study of large, freely evolving, protein populations over long time courses.

In contrast, long evolutionary trajectory experiments have been successfully executed for populations of whole organisms and RNA. Seminal work by Lenski and others studying the evolution of whole organisms through continuous culture^{16–20} have elucidated some of the determinants of organismal evolutionary outcomes including the effects of population size, the role of epistasis, and the importance of evolvability^{21–24}. Additionally, bacteriophages have been used as a relatively minimal, rapidly reproducing system for experimental evolution at the whole-genome level^{25,26}. Organismal evolution can be difficult to dissect at a molecular level, however, as mutations typically occur not only in genes of interest but also throughout the host genome. Fitness gains *in vivo* are therefore frequently influenced by complex sets of mutations, confounding the elucidation of the molecular determinants of fitness gains²⁷ at the protein level. Phage display and related techniques can constrain evolution to a small set of genes of interest, but these methods, being more akin to screening, are generally too cumbersome to support many (e.g., dozens or hundreds of) generations of evolution²⁸.

RNA continuous evolution methods have enabled long evolutionary trajectory experiments on both RNA genomes^{24, 25} and catalytic RNAs^{29–32}. These elegant experiments demonstrate the power and potential of continuous evolution methods applied over long time courses. In both cases, the development of methodology and infrastructure allowing for continuous evolution enabled the study of long evolutionary trajectories. However, these methodologies rely on fundamental features of RNA replication and have not been applied to proteins.

Long evolutionary trajectories have not been studied on the single protein level in part due to a lack of a methodology capable of supporting protein continuous evolution. Recently we developed phage-assisted continuous evolution (PACE), a method for the continuous directed evolution of proteins³³ that performs the selection, replication, and mutation of genes of interest continuously without human intervention. PACE enables up to ~40 theoretical rounds of evolution to take place every 24 h³³. The PACE system selectively propagates selection phage (SP) that encode evolving proteins in a continuously diluted fixed-volume vessel (a “lagoon”) by linking the activity of SP-encoded proteins to the production of an essential phage protein, pIII, encoded by gene III. The *E. coli* cells contain an accessory plasmid (AP) that is the only source of gene III in the system (Figure 1). Phage possessing active SP-encoded proteins are capable of generating infectious progeny while phage possessing inactive SP-encoded proteins are not. Importantly, due to the rate of the continuous dilution, the host *E. coli* do not have sufficient time to divide before they exit the lagoon, preventing their evolution and ensuring that only the phage-encoded genes evolve.

The nature of PACE enables mutations to accumulate exclusively in the phage genome. Previous *in vivo* evolution studies have used mutator *E. coli* strains^{34, 35}, which introduce mutations throughout both the gene of interest and the *E. coli* genome, complicating the interpretation of fitness gains and necessitating human intervention between rounds. In PACE, the host *E. coli* cells possess an arabinose-inducible mutagenesis plasmid (MP) that is only induced in the lagoon. Like traditional mutator strains, mutations are distributed across the gene of interest and the host. However, unlike traditional mutator strains, mutations persist in the phage genome and not in the *E. coli* host since the average residence time of the *E. coli* cells in the lagoon is insufficient to enable cell division.

The uncoupling of gene-of-interest evolution from host genome evolution during PACE enables the study of large gene populations over hundreds of rounds of evolution in parallel replicates with minimal human intervention. Moreover, the selection conditions of the gene of interest can be carefully controlled with minimal concern for the impact on cell survival or cell evolution. PACE can therefore serve as an experimental platform for studying the determinants of protein evolution outcomes over long evolutionary trajectories.

In this work we integrated phage-assisted continuous evolution (PACE)³³ and high-throughput DNA sequencing to study the effects of mutation rate and selection stringency on evolving protein populations over long evolutionary trajectories that would be difficult or impractical to implement using conventional directed evolution methods. We observed that specific combinations of mutation rate and selection stringency reproducibly resulted in differences in evolutionary outcomes, including mutational “signatures” and non-signature mutations that map to specific regions of protein sequence. Our findings illuminate key determinants of protein evolutionary outcomes and suggest hypotheses that inform both the design of future protein evolution experiments and the interpretation of natural protein evolution.

MATERIALS AND METHODS

General Methods

All PCRs were performed with Hot Start Phusion II polymerase (Thermo Scientific). Water was purified using a MilliQ water purification system (Millipore, Billerica MA). All vectors were constructed by isothermal assembly cloning³⁶ (*i.e.* Gibson assembly). Single point mutants and reversions were generated using the QuikChange II Site-Directed Mutagenesis Kit (Agilent). All DNA cloning was performed with NEB Turbo cells (New England Biolabs). Plaque assays and PACE experiments were performed using *E. coli* S109 cells derived from DH10B as previously described.³³ Luciferase assays were performed in NEB 10-beta cells (New England Biolabs) as described in Supporting Information.

Phage pre-optimization

To minimize the potential fitness advantages of mutations to the phage genome, a previously described VCM13 helper phage with T7 RNAP (HP-T7RNAP A)³³ was pre-optimized by PACE. HP-T7RNAP A was continuously propagated for 6 days with arabinose induction at a 2.0 volume/h dilution rate using a high-copy AP containing gene III under control of a T7 promoter. Wild-type T7 RNA polymerase (T7 RNAP) was then subcloned into a randomly chosen phage backbone clone from this pre-optimization selection and sequenced to ensure correct cloning of the T7 RNAP gene. The resulting SP (SP T7 RNAP wt) was used as the starting point for all PACE experiments.

Phage-Assisted Continuous Evolution (PACE)

The turbidostat, lagoons, media, and general PACE set-up were set up as previously described.³³ Lagoons were 40 mL with a flow rate of 2.0 volumes/h. Lagoon samples were collected at 6, 12, 24, 30, 36, 48, 54, 60, 72, 78, 84, 96 h. Each lagoon was inoculated with 5×10^4 pfu SP T7 RNAP wt (see ‘phage pre-optimization’ above) and propagated continuously for 48 h on AP-T7/T3. To begin the second 48 h of selection (on AP-T3), 40 μ L of lagoon sample from 48 h was used to reinitiate PACE. Each lagoon contained $\sim 10^9$ to 10^{10} phage after 48 h, corresponding to re-initiation with a population size of $\sim 10^6$ to 10^7 phage per lagoon. This large phage population was used to minimize imposing a bottleneck in the evolution between the hybrid promoter and the final T3 promoter while still enabling further experiments with the phage.

Samples used for re-selection (the sample from the lagoon that washed out and the samples used for the low-then-high stringency selection) that were greater than one month old were revived using the following procedure: 40 μ L of lagoon isolate was added to 500 μ L of fresh cells ($OD_{600} = 0.4$), incubated at 37 °C for 30 min, then added directly to a lagoon to initiate PACE.

High-throughput sequencing data analysis

A custom MATLAB script (available upon request) was used to align HTS sequencing reads to the wild-type sequence and count the nucleotide and amino acid positions from which the experimental sample deviates from the wild-type sequence. We observed an error rate that varies as a function of nucleotide position; importantly, the error rate is highly reproducible from multiple sequencing runs and sample preparations. We sequenced multiple, independently prepared samples (over multiple sequencer runs) of the wild-type gene, and used the error rate of these samples as a ‘baseline’ for future experiments. This yielded both an average error rate and a standard deviation for the error of wild-type sequencing for each nucleotide and amino acid position in the gene (Figure S1).

To compensate for systemic sample preparation and sequencing errors, the observed fraction of mutations at each nucleotide or amino acid position of the wild-type T7-RNAP reference gene was subtracted from the fraction of mutations in a given experimental sample to result in the “corrected fraction mutated”. Mutations were defined as amino acid positions with a corrected fraction mutated that is both $\geq 2.5\%$ and at least five standard deviations higher than the corrected fraction mutation of the wild-type reference sequence. Extensive controls demonstrating the validity of this sequencing methodology are detailed in the Supporting Information (see Supplementary Results, Figures S1 through S4 and Table S1).

Additional Methods are provided in the Supporting Information.

RESULTS

Experimental Design

T7 RNA polymerase (T7 RNAP) is a single-subunit RNA polymerase that recognizes the native T7 promoter with a high degree of specificity³⁷. We used PACE to evolve T7 RNAP to recognize the T3 promoter^{38, 39} (Figure 2A), which is not natively recognized by T7 RNAP, under four distinct selection conditions, each in four-fold replicate: high stringency, high mutagenesis; high stringency, low mutagenesis; low stringency, high mutagenesis; and low stringency, low mutagenesis. We controlled selection stringency by modulating the copy number of the AP, which modulates the concentration of substrate DNA within the cell. We previously demonstrated that infectivity of phage progeny, and thus selection stringency, can be modulated by changing the copy number of the AP³³. The high-copy AP (pMB1 Δ *rop* origin) is present in approximately 20–60 copies per cell⁴⁰ and corresponds to “low stringency” selection conditions, while the low-copy AP (SC101 origin) is present in approximately five copies per cell⁴¹ and corresponds to “high stringency” conditions. The mutagenesis rate was modulated using an inducible mutagenesis plasmid (MP), which enhances the mutation rate of propagating SP by ~ 100 -fold (Figure 2B)³³. All high-mutagenesis PACE lagoons contained 1% arabinose, which increases the mutation rate of the phage produced by approximately 100-fold, sufficient to generate all possible double mutants of T7 RNAP shortly after the gene enters the lagoon. The low-mutagenesis lagoons received an equivalent volume of water, and therefore relied on the basal mutation rate of DNA replication ($\sim 5 \times 10^{-7}$ per nucleotide per generation⁴²) to generate diversity.

All selections began by seeding each lagoon with 5×10^4 pfu SP encoding wild-type (wt) T7 RNAP. Because wild-type phage do not propagate on host cells containing the T3 promoter,

the lagoons were continuously evolved for 48 h on a hybrid T7/T3 promoter (AP-hybrid)³³ that served as an evolutionary stepping-stone to T3 promoter recognition. A sample from each lagoon was then diluted into a fresh lagoon receiving host cells harboring AP-T3 and continuously evolved for another 48 h (96 h total). Phage surviving 96 h of PACE in each lagoon will have undergone an average of ~100 theoretical rounds of evolution³³, calculated based on the theoretical time for an average phage life cycle during PACE, and survived a ~10⁸⁴-fold net dilution.

Genetic evidence for positive selection

To quantitatively analyze population genotypes, we subjected lagoon samples to high-throughput DNA sequencing (HTS). We experimentally demonstrated that HTS could reliably detect mutations present at ~2.5% frequency in each population (Supplementary Results, Figures S1 through S4, Table S1 and S5). Across all lagoons, 153 instances of significantly mutated nucleotide positions were observed; of these, 101 represent coding mutations while 52 represent silent mutations. Of the 101 coding mutations, 32 are observed in more than one lagoon (32%), while only one of the noncoding mutations is observed in multiple lagoons (2%).

The 101 coding mutations result in mutations at 97 of the 883 amino acids of T7 RNAP, representing 11% of the total amino acids of the protein. Among these are a number of mutations that have been previously described to be important for substrate broadening, such as E222K³⁸, to serve as a specificity determinant for T3 promoter recognition such as N748D³⁹, or have been identified in previous work, such as G542V⁴³. Collectively, the strong enrichment of coding mutations over noncoding mutations, the recurrent nature of these mutations, and the observation of known beneficial mutations provide compelling evidence of positive evolution.

All four selection conditions evolve T3 promoter recognition activity

SPs encoding wt T7 RNAP do not form plaques on host cells containing either low or high stringency AP-T3. In contrast, 15 of the 16 lagoons at 96 h contained phage that formed plaques on AP-T3 of their respective stringency. Although all 16 lagoons yielded phage that were active on the T7/T3 hybrid promoter at the conclusion of the T7/T3 hybrid selection (48 h time point), one lagoon repeatedly failed to yield T3-active phage (high stringency, low mutagenesis lagoon #1) at the end of the T3 selection, likely due to its distinct genetic composition following T7/T3 hybrid evolution (see below).

We assayed the activity of ten or more RNAP genes from each of the fifteen active 96-hour lagoons. Although wt T7 RNAP showed no detectable activity on the T3 promoter (< ~1%), the average lagoon activity from all 15 active lagoons at 96 h exhibited activities on the T3 promoter of ~11% of the activity of wt T7 RNAP on the T7 promoter, which we define as 100% (Figure 3). Notably, RNAP variants evolved in the high-stringency lagoons showed an average T3 promoter activity of 215%, whereas the low-stringency lagoons evolved an average T3 promoter activity of 43%. These results indicate that evolved activity levels were strongly dependent on selection stringency.

Potential explanation for phage washout of high stringency, low mutagenesis lagoon #1

To test if the inability of high stringency, low mutagenesis lagoon #1 to survive the final 48 h of selection on AP-T3 was a stochastic occurrence or instead reflected a property of this lagoon's population after 48 h, we repeated the final 48 h of T3 selection for this lagoon in duplicate. Once again no active phage were observed in any replicate after 48 h of high stringency, low mutagenesis selection on AP-T3, indicating that the enzymes at the end of

the 48-hour T7/T3 hybrid selection in this lagoon were not capable of evolving sufficient activity on the T3 promoter.

To begin to understand the inability of one of the lagoons to complete the selection at low mutagenesis, high stringency, we performed HTS on samples from all of the lagoons at 48 h using the methods described above. When we compared genetic data from 48-h samples to genetic data from 96-h samples, we noticed that mutation at E222 varied significantly between 48 and 96 h (Figure S5). At 48 h, a wide range of mutations was present at E222 while at 96 h only E222K and E222Q were observed. Thus, several of the mutations present at 48 h are de-enriched upon selection for activity on T3. One of those de-enriched mutations, E222G, is present in 100% of the population of the lagoon that did not complete the selection. Importantly, while three of the eight high-stringency populations possess E222G as a measurable subpopulation at 48 h, zero possesses E222G at 96 h.

We hypothesized that E222G may be unable to achieve high activity in the context of the other highly enriched mutation of the high stringency, low mutagenesis conditions, N748D. To test this hypothesis, we biochemically reconstructed the three possible combinations of E222K/Q/G and N748D and assayed their activity. We observed that E222K/N748D (the most common combination) is the most active double mutant on T3, while E222G/N748D was the least active double mutant on T3 (Figure S6). While E222G/N748D is more active than N748D alone, it is still less active than the other double mutants, and may not possess sufficient activity to propagate. This could be an example of negative sign epistasis, in which mutations that are individually beneficial become deleterious in combination, which have been previously identified¹² and may have created evolutionary dead-ends. Other possible explanations for the relationship between E222G and phage washout include increased polymerase promiscuity or decreased protein stability, which are beyond the scope of this study. We also note that while our observations of E222G's abundance and effects on activity are suggestive, other, less obvious genetic differences might also underlie the inability of phage in this lagoon to propagate.

Notably, there is significantly more amino acid variation at E222 at 48 h among high mutagenesis, high stringency lagoons relative to low mutagenesis, high stringency lagoons. While several other lagoons contained a measurable subpopulation containing E222G at 48 h, in each case the E222G mutation was not present at 96 h (Figure S5). Unlike the lagoon that did not yield viable phage at 96 h, each of the other lagoons that contained E222G at 48 h also contained a subpopulation containing either E222K or E222Q, which, upon selection on T3, overtook the population. These findings suggest that, particularly under low mutagenesis conditions, different replicates can lead to dramatically different outcomes, including the inability to complete an evolutionary trajectory.

Selection conditions determine genotypic outcome

Populations of RNAP genes obtained through HTS data were characterized using three metrics that reflect different aspects of genetic diversity and one new metric that we developed to reflect evolutionary divergence and reproducibility. M is the total number of unique mutations ($\geq 2.5\%$ frequency in the population) present in the set of four lagoon replicates for a given selection condition, $\langle M \rangle$ is the average number of mutations in each of the four lagoon replicates across a given selection condition, and ISI (inverse Simpson index, averaged across loci) measures the average genetic diversity over all populations⁴⁴. To measure the similarity of evolved populations we developed a new metric, F_{div} (Supporting Information), which is based on F_{ST} ⁴⁵ but is modified to reflect the divergence between populations relative to the divergence from an ancestral starting gene. For the purpose of detecting divergent or parallel evolution from a known ancestor, F_{div} has the advantage that, if a mutation at a particular locus fixes in two separate populations, this will

decrease the value of F_{div} — indicating greater parallelism— relative to the case that neither population has a mutation at this locus. In contrast, F_{ST} treats these two cases as equivalent. As with F_{ST} , compared populations are increasingly divergent and dissimilar as F_{div} approaches 1.

The highest values for all three diversity indices (Table 1) were observed for the low stringency, high mutagenesis condition, and the lowest values were observed for the high stringency, low mutagenesis condition, indicating that higher mutation rates and lower selection stringency increase the mutational diversity of evolved populations, as expected. We calculated F_{div} for each selection condition and for each possible pair of selection conditions to compare the similarity between sibling lagoons (those evolved under identical conditions) with the similarity of non-sibling lagoons (Figure 4 and Table S2). Among the ten F_{div} values, the three smallest F_{div} values were from populations within a given selection condition (low stringency, high mutagenesis $F_{\text{div}}=0.29$; high stringency, high mutagenesis $F_{\text{div}}=0.27$; high stringency, low mutagenesis $F_{\text{div}}=0.29$). These results indicate that the most similar evolutionary outcomes occurred within one set of selection conditions, implying that selection condition similarity is the primary determinant of the relatedness of evolutionary outcomes. Induction of mutagenesis under low or high stringency selection increased all three metrics of diversity compared to low mutagenesis populations, but surprisingly resulted in lower F_{div} values under low stringency. This finding indicates that increased mutagenesis results in increased evolutionary convergence of populations relative to the genetic distance separating them from their common ancestor.

Selection conditions generate distinct mutational signatures

With the exception of mutations at E222, most high-abundance mutations—defined as those that appear in at least 50% of a lagoon’s population— are observed in multiple sibling lagoons, but occur only rarely within lagoons subjected to other selection conditions. The similarity between the high-abundance mutations in sibling lagoons creates a common set of convergent mutations— a “mutational signature”— unique for that selection condition (Figure 5). For example, the populations evolved under high stringency, high mutagenesis conditions contain E222K and N748D at ~100% abundance and Y178H/C/D at 57% abundance. This particular pattern of mutations is absent in lagoons subjected to the other three selection conditions. Genes evolved under high stringency, low mutagenesis conditions possess a related, though distinct signature that lacks mutation at Y178 (100% mutation of E222K/Q and N748D). The low stringency, high mutagenesis populations evolved a different signature of E222K, E775K, and either G542V or V574A, which appear to be negatively epistatic (Supporting Information, Figures S7 and S8). Each of these three mutational signatures is common to sibling lagoons, but was not found in any of the 12 non-sibling lagoons. As suggested by its high F_{div} value ($F_{\text{div}} = 0.90$, the highest observed value), no signature mutations emerged from the low stringency, low mutagenesis selection conditions, even though enzymes surviving these conditions evolved significant T3 promoter activity.

Non-signature mutations converge on specific amino acid regions

Although the reproducibility of the mutational signatures is striking, the vast majority (91%) of mutations are not found in a mutational signature. These non-signature mutations (NSMs) are present at < 50% abundance and many (30%) evolved independently in multiple lagoons. While these NSMs did not strongly converge on specific amino acids, we wondered if these mutations converged on particular regions of primary amino acid sequence. To test this possibility, we computationally generated 1,000 sets of simulated protein sequences with an identical total mutational frequency as the fifteen 96-hour lagoons but with these mutations randomly distributed throughout the simulated proteins. We then tested if the

experimentally observed NSMs are more frequently clustered within 10-amino acid segments covering the entire protein than in the simulated sequences (Supporting Information).

We observed that the experimental data set has a greater number of 10-amino acid segments containing five or more NSMs than any of the 1,000 simulated sequences ($p < 0.001$), suggesting that the observed mutations are more tightly grouped than would be expected by random chance. There are nine different overlapping protein segments (“clusters”) that contain five or more mutations (Table S3). Based on the crystal structure of the initiation phase T7 RNAP^{46, 47}, which is thought to be the most important for promoter recognition and clearance, four of these windows (amino acids 121–136, 230–247, 379–394, and 763–779) are predicted to make direct contact with either the DNA substrate, the nascent RNA transcript, or the incoming NTP. Perhaps more surprisingly, five of these amino acid regions (56–76, 153–184, 453–469, 595–607, and 672–697) do not make direct contacts with substrate in the initiation phase T7 RNAP structure, but all but one of these regions (453–469) makes DNA or RNA contacts in the elongation structure⁴⁸, suggesting optimization of these regions may improve polymerase parameters aside from promoter clearance.

To analyze whether NSM clusters are specific to selection conditions, we repeated the simulations described above for each selection condition separately (Supporting Information). Although we did not identify clusters that were specific to either of the two low-mutagenesis selections, three clusters are specific to the high stringency, high mutagenesis selection conditions and three different clusters are specific to the low stringency, high mutagenesis conditions. Interestingly, in five of these six clusters, we observed significantly enriched mutations as a function of selection stringency (Figure 6A,B). These results suggest that NSMs tend to converge on specific regions of primary sequence that are dependent on selection stringency, and therefore experimentally demonstrate that selection conditions can cause convergence on mutations in distinct regions of primary sequence.

Stringency defines distinct mutational paths through fitness landscape

The differences in both the mutational signature and the NSM clusters in the populations selected at high and low stringency suggest that the low-stringency solutions may not be on the same mutational path⁴⁹ as their high-stringency counterparts. These results are consistent with a model in which at least two distinct mutational paths through the fitness landscape, one associated with high activity and one associated with lower activity are accessed in a stringency-dependent manner. To test this hypothesis, lagoon samples from all four of the 96-hour low-stringency, high-mutagenesis lagoon replicates were separately evolved on AP-T3 for an additional 48 h under high-stringency, high-mutagenesis conditions (“low-then-high stringency”). If the resulting populations converge on the genotypes of the populations evolved directly under high stringency, high mutagenesis conditions, then these results would suggest that both conditions guided the populations through similar mutational fates. Alternatively, if the resulting genotypes remain segregated based on selection stringency history, this result would suggest that the populations followed different evolutionary paths to reach a final high-activity endpoint.

While only one of the four low-stringency, high-mutagenesis lagoons had average activity on the T3 promoter greater than 35% at 96 h, all four of the lagoons evolved at least 115% average activity after 48 additional hours of high stringency selection, very similar to the activity levels of the original 96-hour high stringency populations (Figure 7A). Despite the similar phenotypes of the high-stringency and the low-then-high stringency populations, the resulting populations still possess strong genotypic differences. The low-then-high stringency populations possess a mutational signature consisting of E222K, G542V/V574A,

and N748D, different than the mutational signature of any other selection condition (Fig. 7B). M , F_{div} , and ISI of the newly evolved populations increased while $\langle M \rangle$ remained the same, indicating an increase in diversity and divergence without an increase in the number of average mutations per lagoon (Table S4). F_{div} between the new population and both low stringency selections increased while F_{div} between the new population and both of the high stringency selections decreased (Figure 7C and Table S4), indicating that the population became more high stringency-like during the course of the selection, but still retained significant low-stringency character. The observation that the low-then-high stringency selection resulted in a different genotype than the high stringency selections suggests that the low-stringency mutational signature arises from a distinct mutational path through the fitness landscape.

Discussion

PACE supports multiple parallel lagoons each containing up to $\sim 10^9$ different mutants in a gene of interest, minimizes population bottlenecks, and enables hundreds of theoretical rounds of evolution in multiple replicates to be performed on a practical time scale (days to weeks). Coupled with HTS, PACE enables population-level studies of protein evolution. The tunable nature of our system allowed us to vary several fundamental parameters of evolution (selection stringency and mutation rate), and examine how each affected phenotypes, genotypes and the reproducibility of those genotypes. Our findings reveal in molecular detail that changing the mutation rate, selection stringency, or the schedule of selection pressures can result in distinct and reproducible population-wide genetic differences. PACE provides an opportunity to directly observe protein evolution in real time on the population level, which could help validate the knowledge gained from previous theoretical models and experimental systems. We note, however, that because the conditions of any laboratory evolution experiment including the ones described here differ from those found in other experiments or in nature, the extent to which these effects will be consistently observed across various protein evolution systems remains to be tested.

Families of related solutions (mutational signatures) reproducibly emerged as a function of selection conditions and changing either mutation rate or selection stringency resulted in a unique mutational signature. Mutations at E222, which were highly enriched under all selection conditions except the low-stringency, low-mutagenesis condition, are known to broaden substrate specificity³⁸. Both high-stringency conditions enriched N748D, a known specificity determinant for T3 promoter recognition³⁹ that is found in native T3 RNA polymerase and that makes direct contacts with the DNA promoter at the T7 RNAP specificity loop⁴⁶. Accessing high T3 promoter activity regimes, as was observed with both high-stringency conditions, may require this direct T3 promoter-RNAP interaction. In further support of this hypothesis, all four populations that went through the low-then-high stringency selection conditions enriched N748D as they obtained high activity on the T3 promoter. While the experiments presented here involved a wider bottleneck than many previous studies (to $\sim 10^7$ phage per lagoon), the role of this decrease in population size on the genotypic outcomes is unknown. Similar PACE experiments can be used in the future to study the effects of different-sized bottlenecks on evolutionary outcomes by seeding the T3 promoter evolution with different amounts of phage from the hybrid promoter lagoons.

The high-stringency, high-mutagenesis condition also enriched mutations at Y178, which makes potential DNA contacts in the elongation complex⁵⁰. This mutation, which was not enriched in the high-stringency, low-mutagenesis condition, may provide additional fitness advantages not directly observed in the transcriptional reporter assay. For example, mutations at Y178 may confer increased stability and therefore increased tolerance of the mutational load experienced by populations under high-mutagenesis conditions^{51, 52}.

Low-stringency, high-mutagenesis conditions resulted in the unique enrichment of E775K, which makes contacts with DNA in the initiation complex⁴⁶, and either G542V, which is located near the DNA substrate and has previously been identified as a mutation associated with broadened ribonucleotide substrate scope⁴³, or V574A. These combinations of mutations resulted in lower average T3 activity than those populations with N748D. Intriguingly, the low-stringency, high-mutagenesis populations scored the highest on all measures of diversity, including the total number of unique mutations, average number of mutations, and inverse Simpson index. Therefore, this unique mutational signature is likely not a result of these populations simply evolving less. Instead, it appears that the combination of G542V/V574A and E775K provides additional benefits under low-stringency, high-mutagenesis conditions, even though that combination has lower measured T3 promoter activity than those populations with N748D. Consistent with this model, E775K was generally de-enriched when the low-stringency, high-mutagenesis populations were evolved under high-stringency conditions. This observation also suggests a potential epistatic interaction between N748D and E775K, even though these residues are not located in close proximity to one another.

Given that T3 promoter recognition activity strong enough to survive the high-stringency selection should also pass the low-stringency selection, it is not clear why the mutational signature in the low-stringency, high-mutagenesis selections reproducibly differs from that of the high-stringency, high-mutagenesis conditions. The signatures obtained in these two conditions not only differ, but have very few residues in common; only E222K is observed in both mutational signatures. These observations suggest that the low-stringency signature is not merely a stepping stone to the high-stringency signature, but instead represents a unique genotype. The high reproducibility of these two mutational signatures argues against stochasticity as the basis for their differences and instead suggests that they may make important contributions to aspects of fitness beyond simply increasing T3 promoter transcription levels.

Induction of mutagenesis resulted in increased diversity, as expected, but also increased the reproducibility of evolution. This effect may arise for a variety of reasons: 1) many possible solutions may satisfy the selection criteria, in which case enhanced mutational sampling may allow more frequent access to a smaller set of superior solutions; 2) increasing mutation rate may implicitly select for the ability to tolerate additional mutations, and satisfying this additional constraint might contribute to the decreased observed divergence; or 3) high mutational loads may result in a high fraction of completely inactivated genes, resulting in narrow bottlenecks of surviving genes and sequence convergence^{53, 54}. The precise molecular underpinnings and potential biological relevance of this observation merit future investigation.

Increasing selection stringency also significantly influenced evolutionary outcomes, resulting in fewer mutations and lower diversity, but more reproducible results. These outcomes are consistent with a smaller set of possible outcomes that satisfy the criteria of the selection and a more constrained fitness landscape.

We also observe non-signature mutations (NSMs) that vary both between and within selection conditions and that converge on regions of primary sequence in a selection condition-dependent manner (“clusters”). It is tempting to speculate that these mutations may be the result of functional redundancy or epistasis with the signature mutations. For example, clustered NSMs might have redundant functions of optimizing positioning or folding of a piece of secondary structure altered by the presence of a beneficial, but not optimal, signature mutation, as has been hypothesized in the “evolutionary Stokes shift” theory⁵⁵. The enrichment of these stabilizing, compensatory mutations are critical under

high mutation rates⁵⁶ and in the acquisition of new functions⁵⁷. This further optimization is consistent with the prevalence of clusters predicted to make substrate contacts in the elongation phase of T7 RNAP, but not in the initiation phase. Regardless of their origin, this significant and curious phenomenon would be difficult to observe without many parallel replicates as well as HTS, suggesting that further application of PACE as an evolutionary model system might elucidate important additional insights into molecular evolution.

These findings have implications for future protein evolution experiments. Our observation that different signature and non-signature mutational solutions evolve in response to changes in mutation rate, selection stringency, and stringency history suggests that performing parallel laboratory evolution experiments under varying mutagenesis levels, stringencies, and stringency schedules may yield a broader diversity of evolved solutions than a more conventional approach of increasing selection stringency of a single population as the number of completed rounds of evolution increases. In addition, since a common goal of some early-stage laboratory protein evolution efforts is to generate a diverse set of modestly active variants prior to a later stage in which these variants are recombined and compete,⁵⁸ our results suggest that low-stringency, low-mutagenesis conditions— conditions that resulted in this study in a broad, consensus-free population of modestly active variants— are well-suited to the early stages of evolution, while increasing stringency or increasing mutagenesis can help drive early-stage populations to high-activity consensus.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by DARPA HR0011-11-2-0003, DARPA N66001-12-C-4207, and the Howard Hughes Medical Institute. A.M.L. was supported by a NIH National Research Service Award Postdoctoral Fellowship (F32GM095028). B.C.D. is a Fellow of the Jane Coffin Childs Memorial Fund for Medical Research. B.A. is supported by the Foundational Questions in Evolutionary Biology initiative of the John Templeton Foundation. D.R.L. is an Investigator with the Howard Hughes Medical Institute. We thank Dr. Kevin Esvelt for helpful discussions.

ABBREVIATIONS

PACE	phage-assisted continuous evolution
HTS	high-throughput sequencing
SP	selection phage
AP	accessory plasmid
MP	mutagenesis plasmid
RNAP	RNA polymerase
wt	wild-type
M	total number of unique mutations
<M>	average number of mutations
ISI	inverse Simpson index
NSM	non-signature mutation

References

1. Vasserot AP, Dickinson CD, Tang Y, Huse WD, Manchester KS, Watkins JD. Optimization of protein therapeutics by directed evolution. *Drug Discov Today*. 2003; 8:118–126. [PubMed: 12568781]
2. Turner NJ. Directed evolution drives the next generation of biocatalysts. *Nat Chem Biol*. 2009; 5:567–573. [PubMed: 19620998]
3. Hancock SM, Rich JR, Caines ME, Strynadka NC, Withers SG. Designer enzymes for glycosphingolipid synthesis by directed evolution. *Nat Chem Biol*. 2009; 5:508–514. [PubMed: 19525967]
4. Gupta RD, Goldsmith M, Ashani Y, Simo Y, Mullokandov G, Bar H, Ben-David M, Leader H, Margalit R, Silman I, Sussman JL, Tawfik DS. Directed evolution of hydrolases for prevention of G-type nerve agent intoxication. *Nat Chem Biol*. 2011; 7:120–125. [PubMed: 21217689]
5. Bornscheuer UT, Huisman GW, Kazlauskas RJ, Lutz S, Moore JC, Robins K. Engineering the third wave of biocatalysis. *Nature*. 2012; 485:185–194. [PubMed: 22575958]
6. Bawazer LA, Izumi M, Kolodin D, Neilson JR, Schwenzer B, Morse DE. Evolutionary selection of enzymatically synthesized semiconductors from biomimetic mineralization vesicles. *Proc Natl Acad Sci U S A*. 2012; 109:E1705–1714. [PubMed: 22679283]
7. Lunzer M, Miller SP, Felsheim R, Dean AM. The biochemical architecture of an ancient adaptive landscape. *Science*. 2005; 310:499–501. [PubMed: 16239478]
8. Weinreich DM, Delaney NF, Depristo MA, Hartl DL. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science*. 2006; 312:111–114. [PubMed: 16601193]
9. Dean AM, Thornton JW. Mechanistic approaches to the study of evolution: the functional synthesis. *Nat Rev Genet*. 2007; 8:675–688. [PubMed: 17703238]
10. O'Maille PE, Malone A, Dellas N, Andes Hess B Jr, Smentek L, Sheehan I, Greenhagen BT, Chappell J, Manning G, Noel JP. Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. *Nat Chem Biol*. 2008; 4:617–623. [PubMed: 18776889]
11. Romero PA, Arnold FH. Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Biol*. 2009; 10:866–876. [PubMed: 19935669]
12. Salverda ML, Dellus E, Gorter FA, Debets AJ, van der Oost J, Hoekstra RF, Tawfik DS, de Visser JA. Initial mutations direct alternative pathways of protein evolution. *PLoS Genet*. 2011; 7:e1001321. [PubMed: 21408208]
13. Gumulya Y, Sanchis J, Reetz MT. Many pathways in laboratory evolution can lead to improved enzymes: how to escape from local minima. *Chembiochem*. 2012; 13:1060–1066. [PubMed: 22522601]
14. Ghadessy FJ, Ong JL, Holliger P. Directed evolution of polymerase function by compartmentalized self-replication. *Proc Natl Acad Sci U S A*. 2001; 98:4552–4557. [PubMed: 11274352]
15. Barlow M, Hall BG. Experimental prediction of the evolution of cefepime resistance from the CMY-2 AmpC beta-lactamase. *Genetics*. 2003; 164:23–29. [PubMed: 12750318]
16. Cooper TF, Rozen DE, Lenski RE. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proc Natl Acad Sci U S A*. 2003; 100:1072–1077. [PubMed: 12538876]
17. Meyer JR, Dobias DT, Weitz JS, Barrick JE, Quick RT, Lenski RE. Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science*. 2012; 335:428–432. [PubMed: 22282803]
18. Elena SF, Lenski RE. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet*. 2003; 4:457–469. [PubMed: 12776215]
19. Elena SF, Agudelo-Romero P, Carrasco P, Codoner FM, Martin S, Torres-Barcelo C, Sanjuan R. Experimental evolution of plant RNA viruses. *Heredity (Edinb)*. 2008; 100:478–483. [PubMed: 18253158]
20. Tenaillon O, Rodriguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, Gaut BS. The molecular diversity of adaptive convergence. *Science*. 2012; 335:457–461. [PubMed: 22282810]

21. Elena SF, Wilke CO, Ofria C, Lenski RE. Effects of population size and mutation rate on the evolution of mutational robustness. *Evolution*. 2007; 61:666–674. [PubMed: 17348929]
22. Zwart MP, Daros JA, Elena SF. One is enough: in vivo effective population size is dose-dependent for a plant RNA virus. *PLoS Pathog*. 2011; 7:e1002122. [PubMed: 21750676]
23. Vignuzzi M, Andino R. Closing the gap: the challenges in converging theoretical, computational, experimental and real-life studies in virus evolution. *Curr Opin Virol*. 2012; 2:515–518. [PubMed: 23025914]
24. Woods RJ, Barrick JE, Cooper TF, Shrestha U, Kauth MR, Lenski RE. Second-order selection for evolvability in a large *Escherichia coli* population. *Science*. 2011; 331:1433–1436. [PubMed: 21415350]
25. Husimi Y. Selection and evolution of bacteriophages in cellstat. *Adv Biophys*. 1989; 25:1–43. [PubMed: 2696338]
26. Bull JJ, Molineux IJ. Predicting evolution from genomics: experimental evolution of bacteriophage T7. *Heredity (Edinb)*. 2008; 100:453–463. [PubMed: 18212807]
27. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature*. 2009; 461:1243–1247. [PubMed: 19838166]
28. Pande J, Szewczyk MM, Grover AK. Phage display: concept, innovations, applications and future. *Biotechnol Adv*. 2010; 28:849–858. [PubMed: 20659548]
29. Breaker RR, Joyce GF. Emergence of a replicating species from an in vitro RNA evolution reaction. *Proc Natl Acad Sci U S A*. 1994; 91:6093–6097. [PubMed: 7517040]
30. Wright MC, Joyce GF. Continuous in vitro evolution of catalytic function. *Science*. 1997; 276:614–617. [PubMed: 9110984]
31. Lincoln TA, Joyce GF. Self-sustained replication of an RNA enzyme. *Science*. 2009; 323:1229–1232. [PubMed: 19131595]
32. Sczepanski JT, Joyce GF. Synthetic evolving systems that implement a user-specified genetic code of arbitrary design. *Chem Biol*. 2012; 19:1324–1332. [PubMed: 23102225]
33. Esvelt KM, Carlson JC, Liu DR. A system for the continuous directed evolution of biomolecules. *Nature*. 2011; 472:499–503. [PubMed: 21478873]
34. Bornscheuer UT, Altenbuchner J, Meyer HH. Directed evolution of an esterase for the stereoselective resolution of a key intermediate in the synthesis of epothilones. *Biotechnol Bioeng*. 1998; 58:554–559. [PubMed: 10099292]
35. Carr R, Alexeeva M, Dawson MJ, Gotor-Fernandez V, Humphrey CE, Turner NJ. Directed evolution of an amine oxidase for the preparative deracemisation of cyclic secondary amines. *Chembiochem*. 2005; 6:637–639. [PubMed: 15719348]
36. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA 3rd, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*. 2009; 6:343–345. [PubMed: 19363495]
37. Summers WC, Siegel RB. Transcription of late phage RNA by T7 RNA polymerase. *Nature*. 1970; 228:1160–1162. [PubMed: 5487240]
38. Ikeda RA, Chang LL, Warshamana GS. Selection and characterization of a mutant T7 RNA polymerase that recognizes an expanded range of T7 promoter-like sequences. *Biochemistry*. 1993; 32:9115–9124. [PubMed: 8369283]
39. Raskin CA, Diaz GA, McAllister WT. T7 RNA polymerase mutants with altered promoter specificities. *Proc Natl Acad Sci U S A*. 1993; 90:3147–3151. [PubMed: 8475053]
40. Twigg AJ, Sherratt D. Trans-complementable copy-number mutants of plasmid ColE1. *Nature*. 1980; 283:216–218. [PubMed: 7350544]
41. Stoker NG, Fairweather NF, Spratt BG. Versatile low-copy-number plasmid vectors for cloning in *Escherichia coli*. *Gene*. 1982; 18:335–341. [PubMed: 6290337]
42. Drake JW. A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci U S A*. 1991; 88:7160–7164. [PubMed: 1831267]
43. Chelliserrykattil J, Ellington AD. Evolution of a T7 RNA polymerase variant that transcribes 2'-O-methyl RNA. *Nat Biotechnol*. 2004; 22:1155–1160. [PubMed: 15300257]

44. Nei M, Chesser RK. Estimation of fixation indices and gene diversities. *Ann Hum Genet.* 1983; 47:253–259. [PubMed: 6614868]
45. Holsinger KE, Weir BS. Genetics in geographically structured populations: defining, estimating and interpreting F(ST). *Nat Rev Genet.* 2009; 10:639–650. [PubMed: 19687804]
46. Cheetham GM, Steitz TA. Structure of a transcribing T7 RNA polymerase initiation complex. *Science.* 1999; 286:2305–2309. [PubMed: 10600732]
47. Steitz TA. The structural changes of T7 RNA polymerase from transcription initiation to elongation. *Curr Opin Struct Biol.* 2009; 19:683–690. [PubMed: 19811903]
48. Yin YW, Steitz TA. Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science.* 2002; 298:1387–1395. [PubMed: 12242451]
49. Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ. Empirical fitness landscapes reveal accessible evolutionary paths. *Nature.* 2007; 445:383–386. [PubMed: 17251971]
50. Tahirov TH, Temiakov D, Anikin M, Patlan V, McAllister WT, Vassilyev DG, Yokoyama S. Structure of a T7 RNA polymerase elongation complex at 2.9 Å resolution. *Nature.* 2002; 420:43–50. [PubMed: 12422209]
51. Tokuriki N, Tawfik DS. Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol.* 2009; 19:596–604. [PubMed: 19765975]
52. Bloom JD, Labthavikul ST, Otey CR, Arnold FH. Protein stability promotes evolvability. *Proc Natl Acad Sci U S A.* 2006; 103:5869–5874. [PubMed: 16581913]
53. Drummond DA, Iverson BL, Georgiou G, Arnold FH. Why high-error-rate random mutagenesis libraries are enriched in functional and improved proteins. *J Mol Biol.* 2005; 350:806–816. [PubMed: 15939434]
54. Bershtein S, Tawfik DS. Ohno's model revisited: measuring the frequency of potentially adaptive mutations under various mutational drifts. *Mol Biol Evol.* 2008; 25:2311–2318. [PubMed: 18687656]
55. Pollock DD, Thiltgen G, Goldstein RA. Amino acid coevolution induces an evolutionary Stokes shift. *Proc Natl Acad Sci U S A.* 2012; 109:E1352–1359. [PubMed: 22547823]
56. Bershtein S, Goldin K, Tawfik DS. Intense neutral drifts yield robust and evolvable consensus proteins. *J Mol Biol.* 2008; 379:1029–1044. [PubMed: 18495157]
57. Wang X, Minasov G, Shoichet BK. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J Mol Biol.* 2002; 320:85–95. [PubMed: 12079336]
58. Jackel C, Kast P, Hilvert D. Protein design by directed evolution. *Annu Rev Biophys.* 2008; 37:153–173. [PubMed: 18573077]

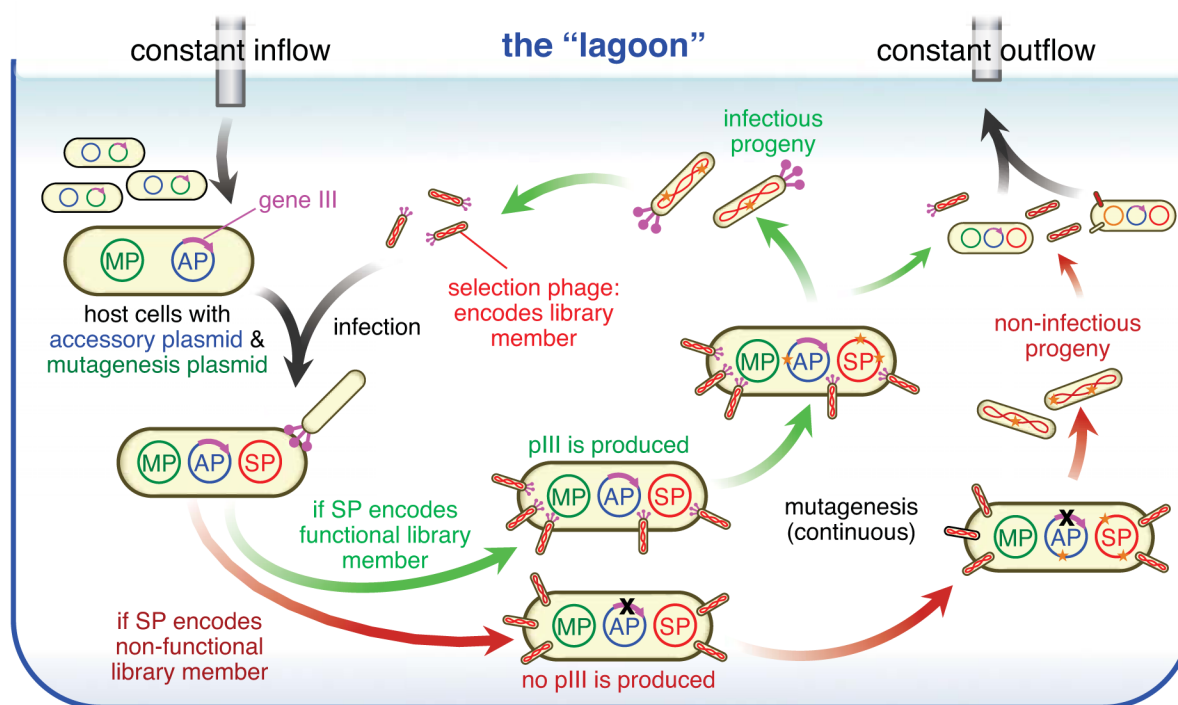
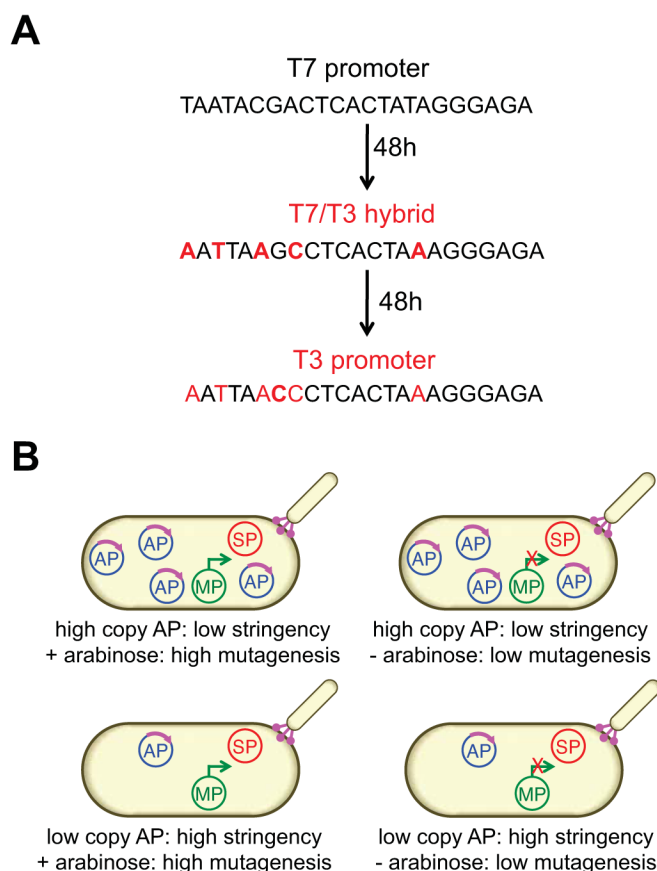


Figure 1.

Schematic overview of phage-assisted continuous evolution (PACE). During PACE selection phase (SP) encoding genes to be evolved propagate in a fixed-volume vessel (a "lagoon"). The activity of corresponding gene products is linked to the production of an essential phage protein, pIII, encoded by gene III. *E. coli* cells, which contain an accessory plasmid (AP) that is the only source of gene III in the system, are continuously pumped into the lagoon. Only phage genomes encoding active proteins of interest induce gene III expression and trigger the production of viable progeny phage. Because the system is constantly diluted, the ability of phage to persist in the system depends directly on their ability to propagate, which in turn depends on the desired activity of the gene of interest.

**Figure 2.**

T7 RNAP promoter evolution as a model to study the effects of mutation rate and selection stringency on protein evolution. (A) DNA sequence of T7 promoter, T7/T3 hybrid promoter, and the final T3 promoter target of the evolution. (B) Schematic of the experimental parameters varied in this study. Stringency was varied by controlling the copy number of the accessory plasmid (AP) and mutagenesis was varied by inducing the expression of mutagenic genes on the mutagenesis plasmid (MP).

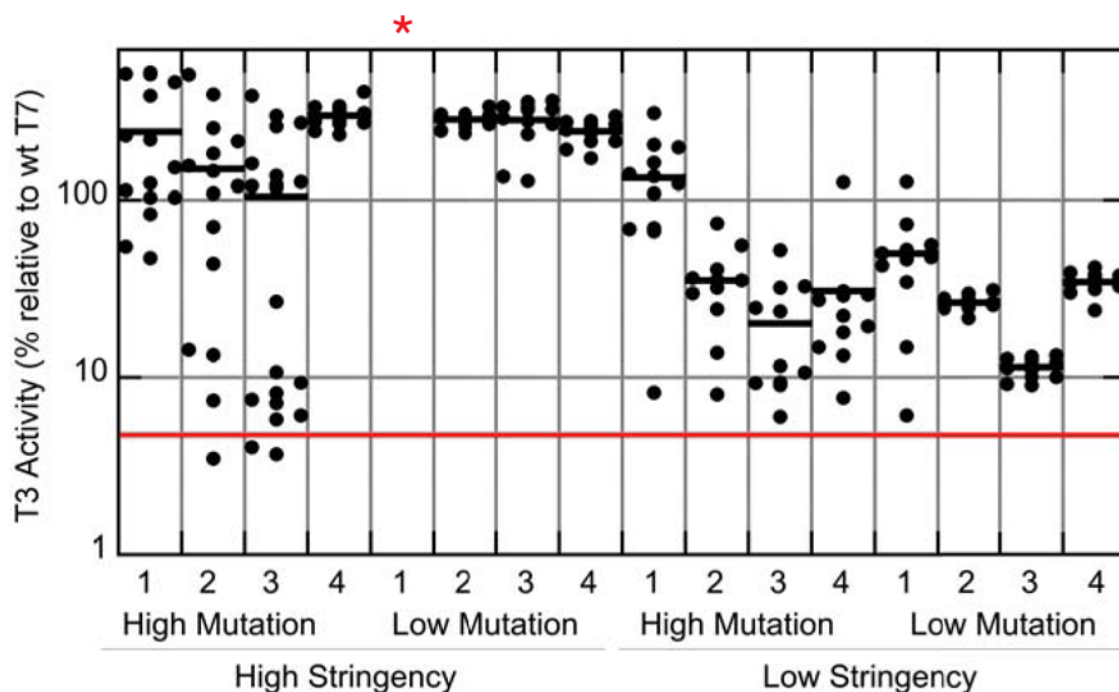


Figure 3.

T3 promoter recognition activity of each lagoon after 96 h of continuous evolution. Each dot represents the relative transcriptional activity of a single randomly chosen clone on the T3 promoter in reporter *E. coli* cells. The black bars represent the average activity of all the assayed clones from one lagoon. The red line represents the endogenous background level of expression of the T3 promoter without any exogenous RNAP. High stringency, low mutation lagoon number 1 resulted in no surviving RNAP genes and is identified with a red asterisk.

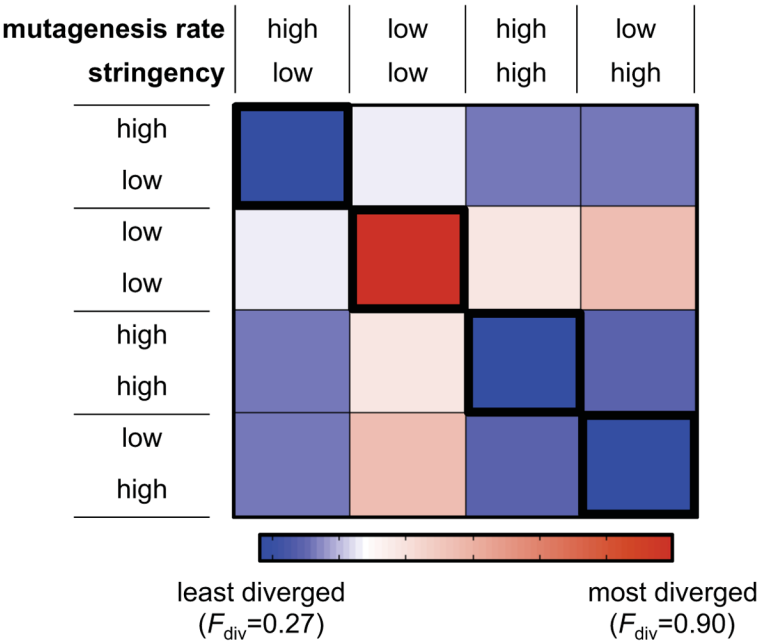


Figure 4. Genotypic outcomes of RNAP evolution under different conditions. F_{div} values at 96 h, calculated for each set of four sibling lagoons and compared to each other set of four non-sibling lagoons.

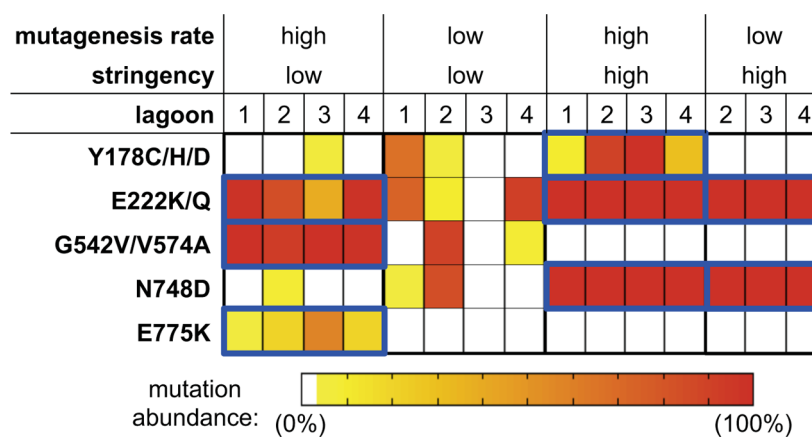


Figure 5. Mutational frequency of positions mutated in at least 50% of any population demonstrating distinct "mutational signatures" that depend on selection conditions.

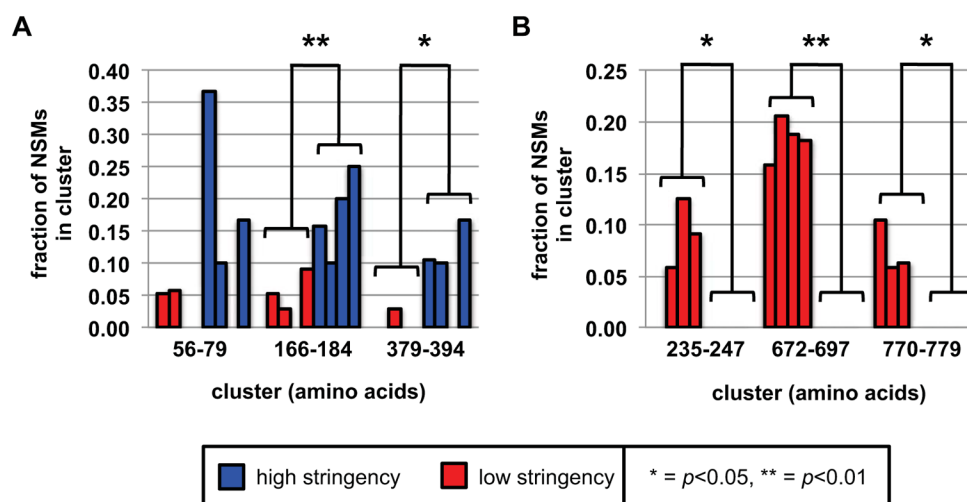


Figure 6.

Clustering of non-signature mutations (NSMs). (A) Fraction of NSMs identified in specific clusters in the high stringency populations. (B) Fraction of NSMs identified in specific clusters in the low stringency populations.

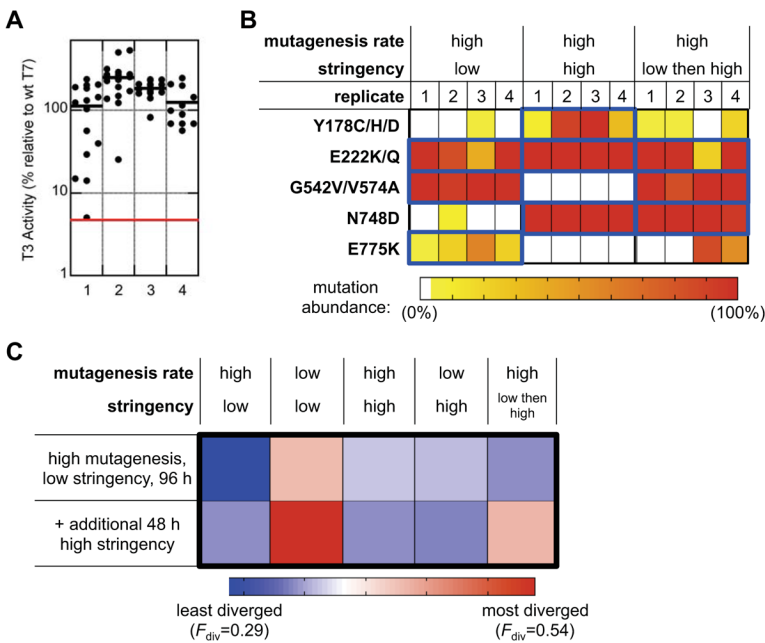


Figure 7. Low-then-high stringency evolution results in distinct genotypes. (A) T3 promoter recognition activity of low stringency, high mutation lagoons after 48 h of additional continuous evolution at high stringency, high mutagenesis. Each dot represents the relative transcriptional activity of a single randomly chosen clone on the T3 promoter in reporter *E. coli* cells. The black bars represent the average activity of all the assayed clones from one lagoon. The red line represents the endogenous level of transcription of the T3 promoter in *E. coli* without any expressed RNAP. (B) Abundance of mutations present in at least 50% of any population, demonstrating that mutational signatures are dependent on selection stringency history. (C) F_{div} values calculated for the low stringency, high mutation lagoons at 96 h and the low stringency, high mutation lagoons after 48 h of additional continuous evolution at high stringency, high mutagenesis compared to all other sets of populations at 96 h.

Table 1

M (the total number of unique mutations present in the set of four lagoon replicates for a given selection condition), $\langle M \rangle$ (the average number of mutations in each of the four lagoon replicates across a given selection condition), ISI (the average genetic diversity over all populations measured by inverse Simpson index, averaged across loci) values for all selections. Error in $\langle M \rangle$ is standard deviation. Error in ISI is calculated according to the description in the Methods.

mutagenesis rate	stringency	M	$\langle M \rangle$	ISI
high	low	64	24±10	1.0068±0.0008
low	low	12	4±2	1.0029±.0008
high	high	41	15±5	1.0045±.0008
low	high	5	3±0	1.0012±.001